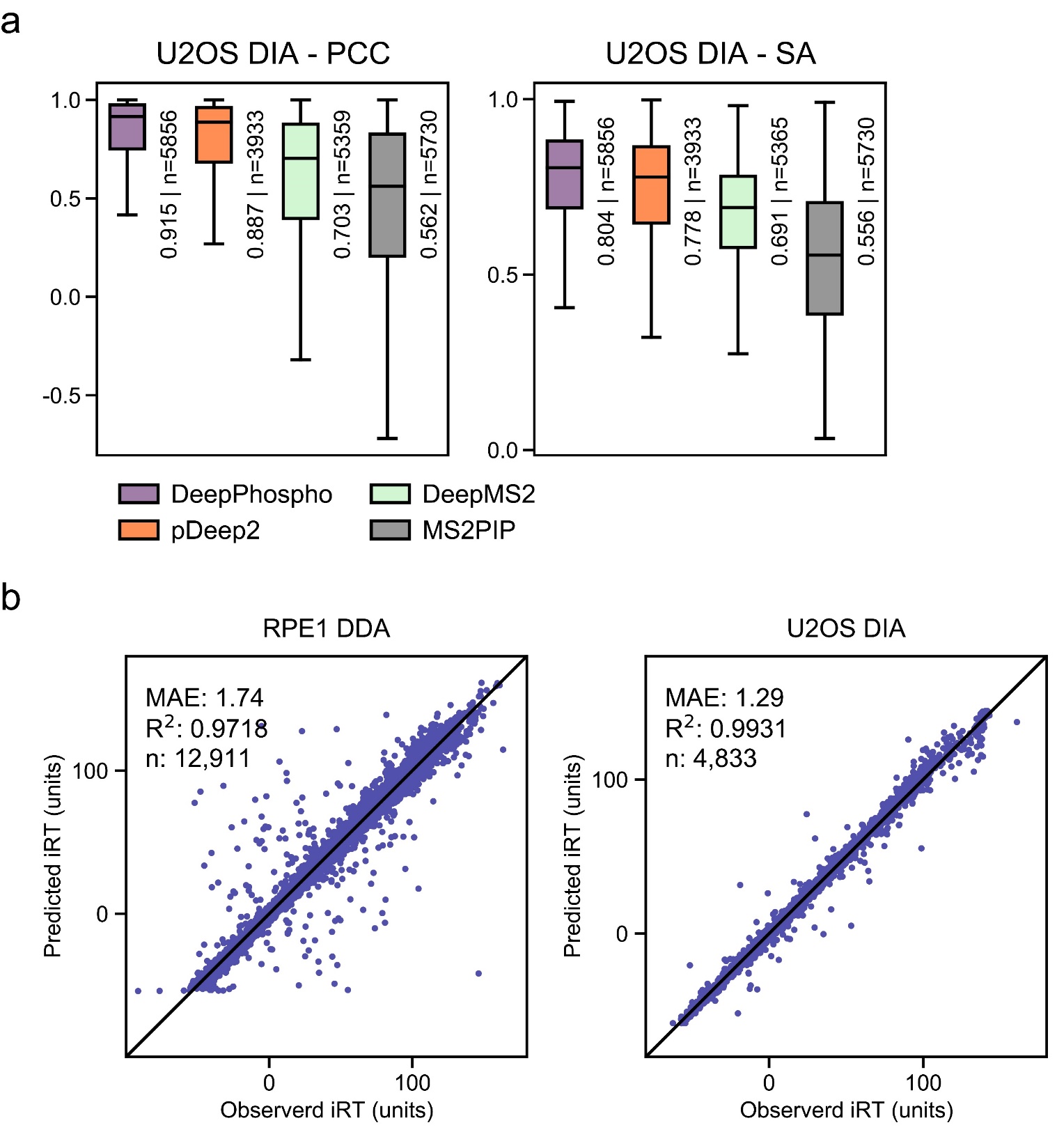


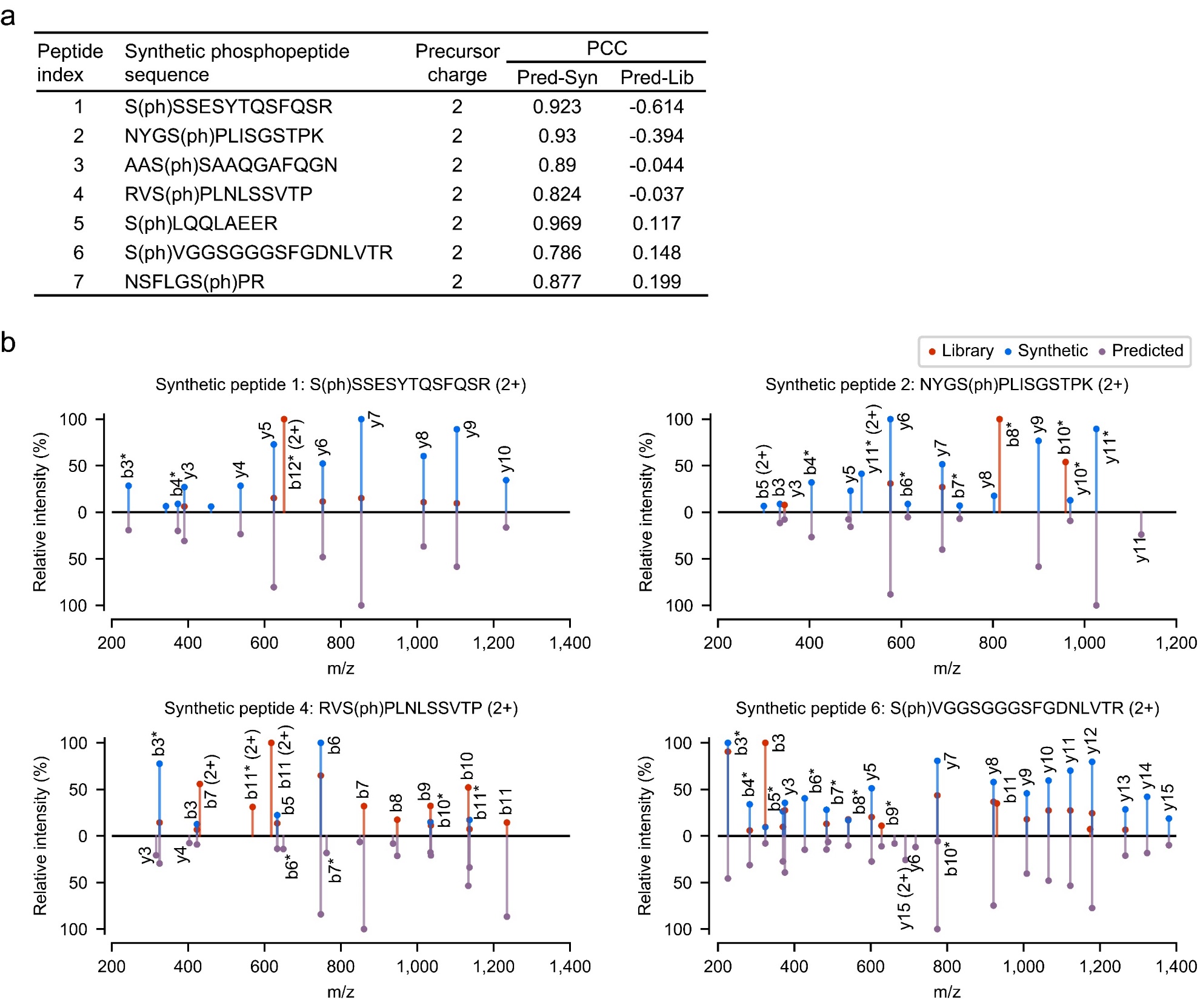
**Supplementary Figure 1 Architecture of DeepPhospho and comparison with other baselines in the ablative study.**

**(a)** Detailed architecture of DeepPhospho. For fragment ion intensity and iRT ­prediction, the embedded features first pass through two stacked bi-directional LSTMs, each of which is followed by a LeakyReLU-Dropout-Linear Layer. After the position encoding is added, the output of biLSTM module is fed into the Transformer module. The first part of each Transformer module is a layer-normalization layer, which is followed by the Multi-Head attention to capture global patterns and a dropout layer to prevent the overfitting. The Transformer module also adopts two skip connections to allow effective model training. **(b)** Evaluation of DeepPhospho and three other baselines based on the distribution of Pearson correlation coefficient (PCC) and spectral contrast angle (SA) calculated between predicted and experimental MSMS spectra from mouse brain DDA and yeast R2P2 DDA datasets. Median PCC and SA are displayed; n is the number of phosphopeptides in the test set. **(c)** Evaluation of DeepPhospho and three other baselines based on the correlation of predicted and experimental iRT values from the yeast R2P2 DDA data. Correlation coefficient of linear regression (R2) and median absolute error (MAE) are displayed.



**Supplementary Figure 2** **Evaluation of DeepPhospho with other datasets.**

**(a)** Evaluation of DeepPhospho and three other models based on the distribution of PCC and SA calculated between predicted and experimental MSMS spectra from the U2OS DIA data. Median PCC and SA are indicated; n is the number of phosphopeptides in the test set. **(b)** Evaluation of DeepPhospho based on the correlation of predicted and experimental iRT values from RPE1 DDA and U2OS DIA datasets. R2 and MAE are indicated.



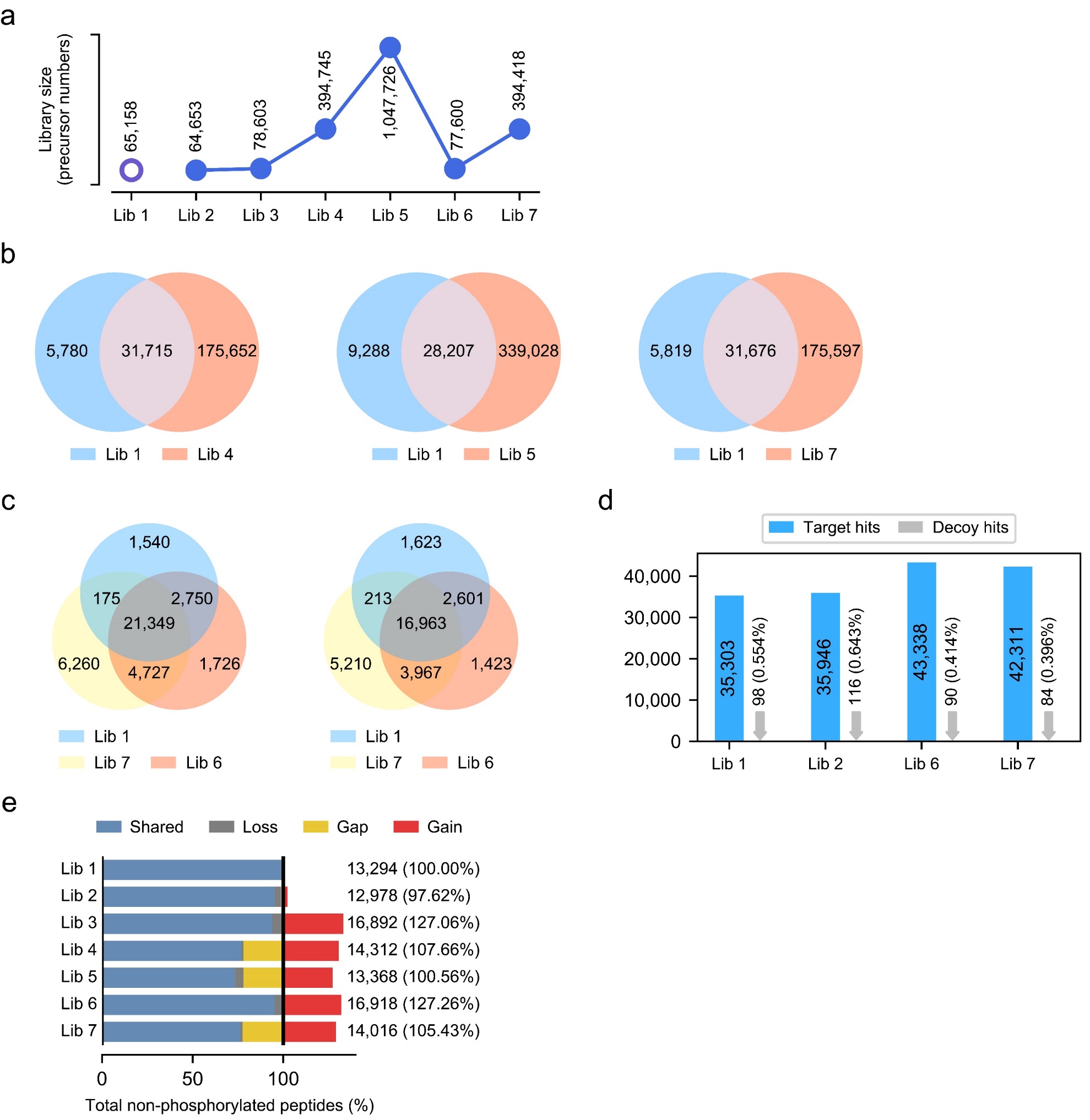
**Supplementary Figure 3** **Spectral similarity analysis for seven** **selected phosphopeptides.**

**(a)** Sequences, charge states and PCC analysis of seven phosphopeptides. Correlation is calculated between the predicted spectra and the high-quality spectra of the synthetic peptide (Pred-Syn), and between the predicted spectra and the DIA library spectra (Pred-Lib). **(b)** Spectra mirror plots for four phosphopeptides not shown in Fig. 2C. Relative fragment ion intensities in the predicted spectra, the DIA library spectra and the synthetic peptide spectra are annotated by purple, orange and blue lines. \* indicates the loss of a phosphate.



**Supplementary Figure 4** **Testing** **21 different conditions in generating the predicted library hPhosPepDB contained in Lib 4 for U2OS DIA data analysis.**

Left table summarizes all 21 combinations of peptide length, precursor and fragment m/z ranges, precursor charge and phosphosite number for the library generation. Right column graphs show the total number of identified phosphopeptides and phosphosites from the U2OS DIA data with each predicted library generated under a specific condition. Condition 20 was selected as the best one for generation of Lib 4 used forU2OS DIA data analysis.



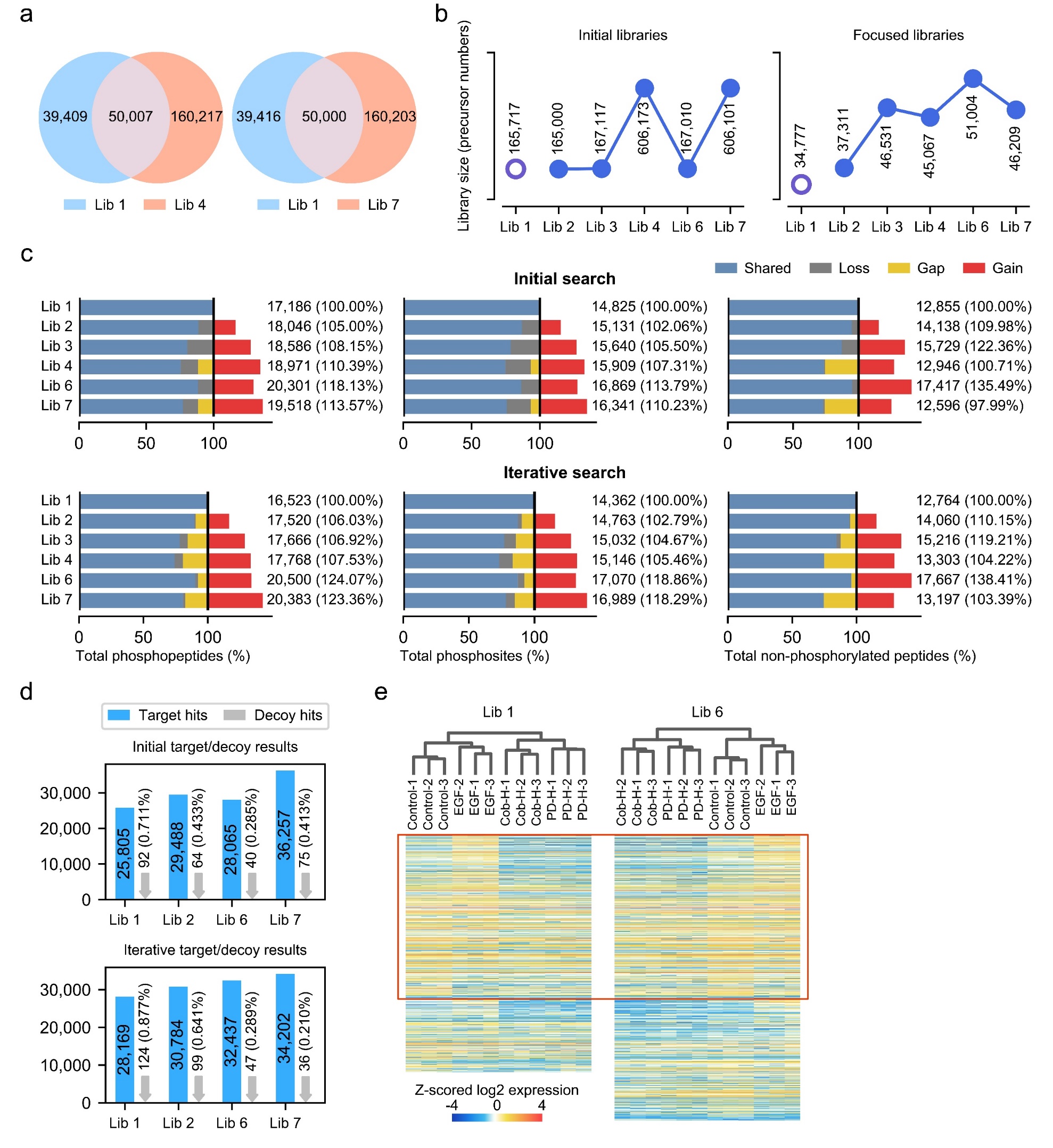
**Supplementary Figure 5** **Comparison of spectral libraries and phosphoproteome profiling results from U2OS DIA data analysis.**

**(a)** Number of total peptide precursors in each generated library. **(b)** Overlapping and unique phosphopeptides present in a DeepPhospho predicted library (Lib 4, Lib 5, Lib 7) *vs* Lib 1. **(c)** Overlapping and unique phosphopeptides (left) or phosphosites (right) identified from U2OS DIA data with Lib 6 and Lib 7 *vs* Lib 1. **(d)** Library-specific FDR assessed using the target-decoy strategy. Number of target peptide IDs and decoy peptide IDs from the U2OS DIA data analysis is shown for each library, with the calculated FDR shown as a percentage. **(e)** Number of non-phosphorylated peptides identified from the U2OS DIA data analysis with each library. Percentage of the total non-phosphorylated peptides number is shown for each predicted library relative to Lib 1. The proportions of shared identifications (IDs), gained IDs, lost IDs and gap IDs yielded by Lib 2 to Lib 7 compared to Lib 1 are indicated in different color. Gap IDs are those present in Lib1 yet absent in the DeepPhospho predicted libraries, thus they cannot be identified with the latter.



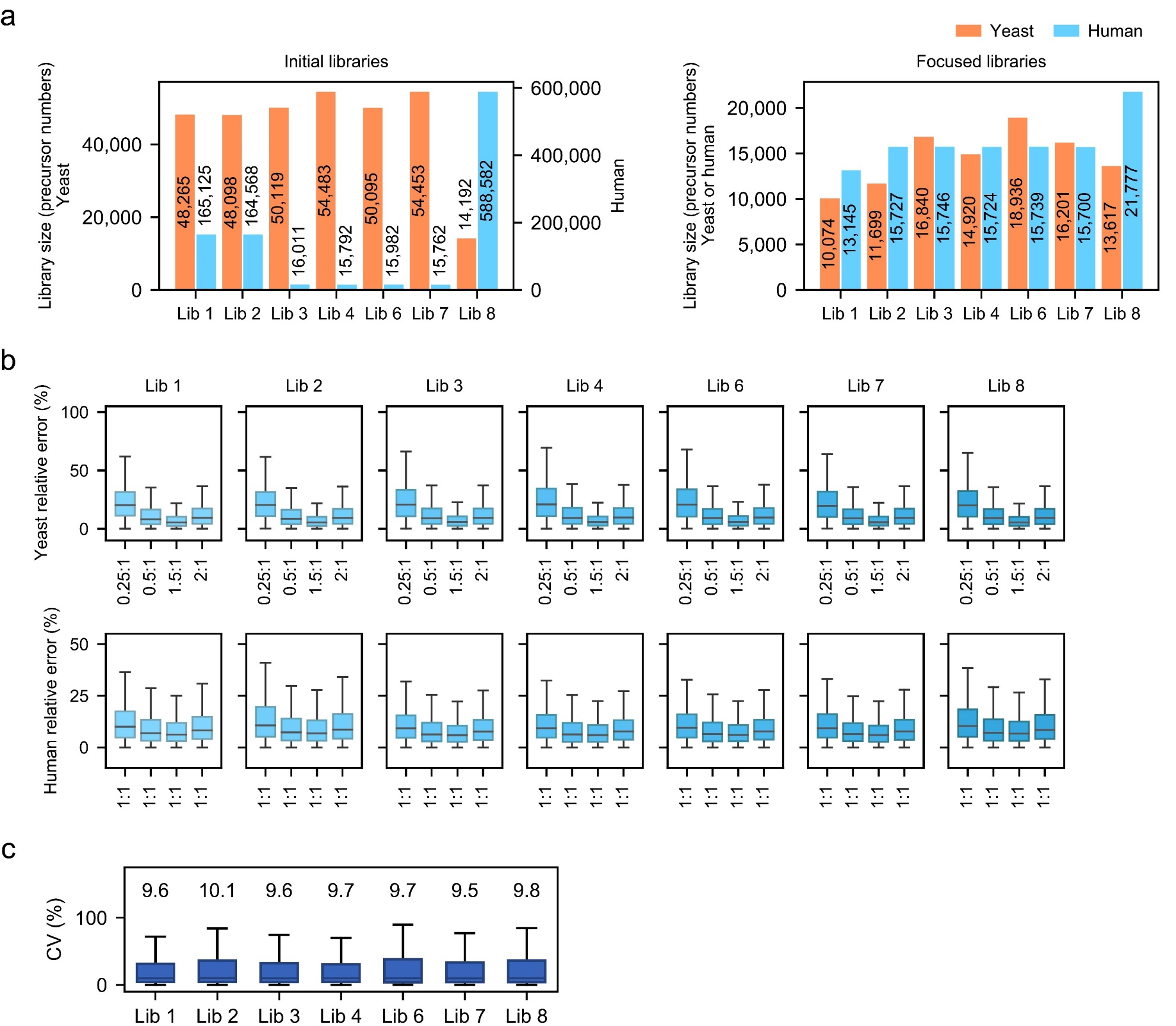
**Supplementary Figure 6** **Testing** **21 different conditions in generating the predicted library hPhosPepDB contained in Lib 4 for RPE1 DIA data analysis.**

Left table summarizes all 21 combinations of peptide length, precursor and fragment m/z ranges, precursor charge and phosphosite number for the library generation. Right column graphs show the total number of identified phosphopeptides and phosphosites from the U2OS DIA data with each predicted library generated under a specific condition. Condition 1 was selected as the best one for generation of Lib 4 used forRPE1 DIA data analysis.



**Supplementary Figure 7 Comparison of spectral libraries and phosphoproteome profiling results from RPE1 DIA data analysis.**

**(a)** Overlapping and unique phosphopeptides identified from RPE1 DIA data with Lib 6 or Lib 7 *vs* Lib 1. **(b)** Number of total peptide precursors in each initial library and the corresponding focused library. **(c)** Number of total phosphopeptides (left), total phosphosites (middle), and total non-phosphorylated peptides (right) identified from RPE1 DIA data with each library in the initial search (upper panel) or in the iterative search (lower panel). Percentage of the total number of identifications is shown for each predicted library relative to Lib 1. The proportions of shared IDs, gained IDs, lost IDs and gap IDs yielded by Lib 2 to Lib 7 compared to Lib 1 are indicated in different color. **(d)** Library-specific FDR assessed using the target-decoy strategy. Number of target peptide IDs and decoy peptide IDs from the RPE1 DIA data analysis is shown for each library, with the calculated FDR shown as a percentage. **(e)** Unsupervised hierarchical clustering of significantly regulated phosphosites yielded at different stimulation conditions with Lib 1 or Lib 6. The red rectangle indicates phosphosites co-identified by two libraries.



**Supplementary Figure 8 Comparison of spectral libraries and phosphoproteome quantification results from DIA data analysis of the two-proteome model.**

**(a)** Number of yeast and human peptide precursors in each initial library and the corresponding focused library. **(b)** Boxplots of relative errors between measured and expected ratios for yeast peptides (upper) and human peptides (lower) from search results with each library. **(c)** % coefficient of variation (CV) of all phosphopeptide quantification with different libraries between 6 replicates at each dilution condition. Median CV% is indicated above the box plot.